True health vs. response styles:
Exploring cross-country differences in self-reported health

Hendrik Jürges

MEA – University of Mannheim and DIW Berlin

15 May 2006

Forthcoming in Health Economics

Mannheim Research Institute for the Economics of Aging
University of Mannheim
L13, 17
D-68131 Mannheim
Germany
Tel: +49-621-181-3519
Fax: +49-621-181-1863
Email: juerges@mea.uni-mannheim.de.

True health vs. response styles:
Exploring cross-country differences in self-reported health

(REVISED VERSION)

Abstract: The aim of this paper is to decompose cross-national differences in self-reported general health into parts explained by differences in "true" health, measured by diagnosed conditions and measurements, and parts explained by cross-cultural differences in response styles. The data used were drawn from the Survey of Health, Ageing and Retirement in Europe 2004 (SHARE), using information from 22,731 individuals aged 50 and over from 10 European countries. Self-rated general health shows large cross-country variations. According to their self-reports, the healthiest respondents live in the Scandinavian countries and the least healthy live in Southern Europe. Counterfactual self-reported health distributions that assume identical response styles in each country show much less variation in self-reports than factual self-reports. Danish and Swedish respondents tend to largely over-rate their health (relative to the average) whereas Germans tend to under-rate their health. If differences in reporting styles are taken into account, cross-country variations in general health are reduced but not eliminated. Failing to account for differences in reporting styles may yield misleading results.

1. Introduction

Self-reports of general health have proved to be useful indicators of an individuals' health, for example as predictors of mortality [1]. However, the comparability of self-reported measures across groups of individuals has been questioned in a number of studies [2,3,4]. One major concern with self-assessed health is that respondents do not perceive the health self-assessment scale given to them as absolute. Individuals with the same true health status may have different reference levels against which they judge their health. For instance, respondents may be likely to report "very poor" health only if they feel they are much less healthy than others of the same sex, age, education, or income. A common finding is that older respondents tend to have a "milder" view of their health, i.e. they tend to rate their health as better than otherwise comparable younger respondents [2,5]. Thus self-reported health of young and old respondents may not be directly comparable, and the observed decline in self-reported health with age may underestimate the decline in true health. In fact, the effect of changing reference levels seems to be so strong that it is taken into account in some formulations of the self-assessed general health question (used e.g. in the BHPS): respondents are explicitly asked to self-report their health relative to other people of their own age.

In cross-cultural studies, there are additional concerns [6]. Respondents from different countries and cultures may not only have different reference levels of health, but response categories may also have different connotations. Self-reported health categories are verbal representations of different health states, which may not mean the same thing to all respondents. For instance, "*excellent*" is a term that is used in everyday parlance in the Anglo-Saxon world, but Germans would often consider "*ausgezeichnet*"

as an ironic exaggeration, in particular if used in the context of health. Another cross-cultural difference in response styles relates to differences in the tendency to choose extreme points of the response scale given to respondents, which will result in more or less variance in reported health. A comparison of self-reported general health across countries has to take such cultural differences into account.

The question is thus whether cross-country differences in self-assessed general health can be taken at face value. If we find that Danes are much more likely to report excellent health than Germans, does that mean that Danes are really that much more healthy than Germans? Or are they simply more likely to report excellent health, even if they have about the same true level of health? Because much can be learned from multi-disciplinary, cross-country comparisons [7], it is of utmost importance to have a good comparable summary measure of the respondents' overall health.

The purpose of this paper is to compute a health measure that is adjusted for possible cross-cultural bias of the kind described above. Conceptually, this is based on a decomposition of differences in self-assessed health into parts that are explained by differences in "objective" health indicators and parts not explained by such differences. The adjusted health measure only contains the explained parts. The basic assumption underlying the analysis is that there is such thing as a "true" and comparable health status. This implies that one must be willing not to accept the respondent's own judgements as absolute [3]. Conceptually, I consider true health as a continuous, latent (i.e., unobservable) variable. When respondents answer survey questions about their health, they assess their true health (possibly with measurement error [8]) and project this value onto the scale provided. Equivalent econometric formulations are the ordered logit or probit models [9]. Differences in language use that affect the relationship

between true health and self-assessed health can be interpreted as differences in the so-called thresholds or cutpoints between adjacent health categories.

The usefulness of this adjustment is demonstrated in a simple policy example studying the cross-national relationship between health care expenditures and self-reported health. The example will show clearly that self-reports can produce spurious results when they are taken at face value.

2. Data description

The data are drawn from Release 1 of the 2004 Survey of Health, Ageing and Retirement in Europe (SHARE). SHARE is modeled closely after the US Health and Retirement Study (HRS) and it is the first European data set to combine extensive information on physical and mental health with information on the income and assets of the older population. The released data contain information on about 22,000 respondents aged 50 and older from 15,000 households in 10 European countries (Sweden, Denmark, Germany, the Netherlands, France, Switzerland, Austria, Italy, Spain, and Greece – future releases will include data from Belgium and Israel). Probability samples have been drawn in each participating country (see [10,11] for details).

Table 1 gives a broad characterisation of the SHARE sample, by country and age group, and sex. Overall sample sizes vary a bit by country. Germany, Sweden, and the Netherlands have the largest samples with about 3,000 observations each. Spain and Italy have slightly smaller samples, with about 2,500 respondents each. Austria, France, Denmark, and Greece have between 1,500 and 2,000 observations and Switzerland has

the smallest sample with about 1,000 respondents. About half of the sample is between 50 and 64 years old and 9.5 percent are aged 80 and over. SHARE also contains information on few respondents who are younger than 50. These so-called "younger spouses", i.e. spouses or partners of age eligible respondents, are omitted from the analyses.

<about here Table 1>

SHARE contains a broad range of different health measures, both of physical and mental health. These include self-reported general health, self reported diagnosed chronic conditions, medication, functional limitations, ADL and IADL limitations, symptoms, mental health as measured by two alternative depression scales (CES-D and Euro-D), physical measurements (hand grip strength and gait speed), self-reported height and weight (to compute BMI) and detailed information on health services utilisation (doctor visits, hospital stays, etc.).

SHARE contains two different 5-point scales for self-rated health, one ranging from "excellent" to "poor", the other from "very good" to "very poor". Each respondent is asked both, in randomised order, at different places of the survey. In this article, I will only report results for the "excellent" to "poor" scale. Overall, this scale seems to work better in the participating countries in the sense that it produces more variance and more symmetric health distributions (translations in each survey language can be found in Table A1 in the Appendix). However, results for the other scale are very similar to those presented in this paper.

Figure 1 shows the age-sex standardised distributions of self-reported general health across the ten countries, i.e. is the health distribution if each country had the same age and sex distribution of individuals aged 50 and over. Countries are ordered by the fraction of respondents who say they are in very good or excellent health. According to their self-reports, the healthiest respondents live in Denmark, Sweden, and Switzerland. The least healthy respondents live in Italy, Spain, and Germany. These findings are remarkable for two reasons. First, compared to other global measures of health such as life expectancy, the difference between the healthiest and the least healthy populations appears to be quite large – probably too large to only reflect differences in true health between countries. Nearly 50% of all Danes aged 50 and over report to be in very good or excellent health, whereas the proportion in Italy is just 19%. But in 2002, the difference between the country with the highest and lowest life expectancy at birth was "only" 3.4 years (Switzerland: 80.6 years, Denmark: 77.2 years [12]). Second, the ranking of the countries by self reported health is at odds with rankings by life expectancy. In terms of life expectancy (at birth), Italy and Spain are among the countries with the healthiest populations (80.2 and 79.7 years, respectively), whereas Denmark has the lowest life expectancy among all 10 countries.

This reversed order somewhat reminds of the difference between the sexes: men usually report better health, but they have lower life expectancy than women [13]. These seemingly contradictory results can be fully explained by differences in the types of chronic conditions men and women usually have. One aim of this paper is to analyse whether the same kind of explanation also applies to cross-national differences in health, or whether there are alternative explanations, such as differential item functioning.

In the present paper, I will use 15 different diagnosed physical conditions (as reported by the respondents), whether ever treated for depression, BMI (derived from self-reported height and weight), grip strength, and walking speed as (quasi-) objective health indicators. I call self-reported diagnosed conditions quasi-objective because they are subjective information on factual matters. Such self-reports can contain some amount of measurement error, usually in the form of under-reporting [14,15]. Below, I will use self-reported diagnosed conditions as explanatory variables in ordered probit regressions. Measurement error could bias their coefficients downwards and thus attenuate their importance relative to more objective measurements. However, in the present paper, I will assume that self-reported conditions reflect true values

Summary statistics for conditions and measurements are reported in Table 2. Again, the numbers are age-sex standardised. Overall, the most prevalent chronic condition is high blood pressure (31.5%), followed by high blood cholesterol (18.4%) and arthritis (18.3%). "Other" conditions have also been reported very often. Although respondents were asked to specify what other condition they meant, this has not yet been coded in the data release used in this paper. Preliminary analyses show that back, hip, and other joint problems are among the most frequent "other" conditions. Cross-country differences in self-reported diagnosed conditions are particularly pronounced for arthritis, with a prevalence of about 30% in Italy, and France, and a prevalence of less than 10% in Sweden and the Netherlands.

<about here Table 2>

Hand grip strength in middle age has been shown to be predictive of the incidence functional limitations, disability and even mortality in old age [16,17]. It is measured using a handheld dynamometer – where respondents are asked to press a lever

as hard as they can. The dynamometer shows grip strength in kg. Of a total of four measurements (two with each hand), I take the largest recorded value. Table 2 shows the proportion of respondents whose hand grip strength – normalised for height and sex – was in the bottom tertile ("low grip strength"), and the proportion of respondents who did not complete the grip strength measurement for health or other reasons ("no grip strength"). Hand grip strength is weakest in Spain and Italy and highest in Austria and Germany. Walking speed is a measure of mobility and functioning of the lower limbs that strongly declines with age (available only for those 75 and over or respondents with self-reported mobility limitations). It is measured by a timed walk over a short distance (2.5m). Two measurements were made, of which I take the fastest. A walking speed of 0.4 m/s or slower is used as the cut-off point for "low walking-speed" [18]. Unsuccessful attempts – independent of the reason – are also coded as having low walking speed. Respondents younger than 75 who were not eligible for the test are coded as having a normal walking speed. Finally, BMI (=weight in kg/squared height in meters) is coded in four categories: <20 (underweight), 20 to 25 (normal weight), 25 to 30 (overweight), and more than 30 (obese). The overall prevalence of obesity is 17%, with the highest proportions found in Greece, Spain, and Austria, and the lowest proportions found in Denmark, Sweden, the Netherlands, and Switzerland.

3. Computing a comparable health index

The first step of the analysis is to construct a 0-to-1 health index, where 0 represents the worst observed health state ("near death") and 1 represents "perfect health". Health states between near death and perfect health are given an index value between 0 and 1. The computed health index will be used as a proxy for true health. The

very basis of any health index is objective information about health problems: diagnosed physical and mental conditions, mental illnesses, and measurements like grip strength, gait speed, and BMI. The absence of any conditions implies perfect health, i.e. an index value of 1. The presence of a condition reduces the health index by some given amount or percentage, the so-called disability weight. The disability weight of each condition or symptom is assumed to be the same for each respondent.

Disability weights are often derived by expert judgements or surveys specialised to elicit health preferences, using time trade-off or standard gambles [19]. Here, I will compute disability weights from within the sample [2,20] by estimating generalised ordered probit regressions of self-reported health on the set of health variables described above. The generalised ordered probit model makes threshold parameters dependent on covariates [21,22]. i.e. health reporting thresholds can vary by respondents' characteristics. Note the double role of self-rated health in this paper. It is first used to construct disability weights for health conditions and later used as the target variable that is adjusted for cross-national comparisons. By construction, the first step is entirely independent from the second. Strictly speaking, it is not even needed. I could as well take disability weights from the literature or even arbitrary weighting schemes (such as equal weights – simply summing up conditions) to form linear combinations of objective health indicators.

Here, the health index is computed as the linear prediction from the ordered probit regression (the latent variable), normalised to 0 for the worst observed health state and 1 for the best observed health state. I deviate from the Cutler-Richardson approach [20], where health is normalised by the range between the lowest and the highest estimated ordered probit threshold value, for two reasons. First, the Cutler-

Richardson approach does not force the predicted health index into the unit interval. Second, I do not need to choose on which set of country-specific thresholds I want base the normalisation.

Disability weights for each condition or impairment are equal to the respective (also normalised) regression parameters. Since the variable on which I base the computation of disability weights is self-reported health itself (and thus potentially subject to cross-cultural bias), I account for country specific reporting styles by modelling the ordered probit thresholds as a function of the country of residence (i.e. I basically have fixed country effects at each threshold). In order to ensure a well-defined probability function for each observation (i.e. that thresholds of higher order are never smaller than thresholds of lower order), threshold equations are specified as (cf. [22]):

$$
\begin{aligned}
\tau_k^1 &= \gamma^1 D_k \\
\tau_k^m &= \tau_k^{m-1} + \exp(\gamma^m D_k), \quad m = 2,\ldots,M-1
\end{aligned}
\tag{1}
$$

where τ_k^m is the m-th threshold for country k, γ^m is a vector of parameters in the m-th threshold equation, and D_k is a vector of country dummies. M is the number of categories of the dependent variable (i.e. five).

While thresholds are allowed to vary across countries, disability weights are constrained to be the same in each country. Note that the cross-national identity of disability weights is the crucial assumption made here to distinguish health from reporting effects. Below, I will relax this assumption and allow for different disability weights across countries when I decompose differences in self-assessed health in parts due to differences in the prevalence of conditions, differences in disability weights and other sources (see section 7).

Table 3 shows the results of the ordered probit estimation and the implied disability weights for each condition or impairment. The first column shows the results from a simple ordered probit regression (with country dummies in the index equation) and the second column contains results from a generalised ordered probit model (with country dummies in the four threshold equations) – the preferred specification. Although a likelihood ratio test rejects the simple ordered probit model at a very high significance level, the parameter estimated in the health index equation do not differ very much. The rejection of the simple ordered probit model mainly comes from the threshold equations. All regression parameters in the index equation have very small standard errors, and are statistically significant. For the sake of simplicity, the specification of the health index equation does not account for co-morbidity. This could be done by a straightforward extension of the model, including interaction effects between different conditions. To check the sensitivity of my results I have repeated my calculations with all first-order interaction effects included. The index equation contained 249 parameters in total. Including the full set of interaction effects did not change the nature of my results.

The third column shows the implied disability weights used in the rest of the paper. They are computed as the regression parameters from the generalised ordered probit model divided by the range of its linear prediction. The highest disability weight is found for Parkinson's disease, followed by stroke, heart attack, and chronic lung disease. "Other" conditions also have a high weight. In a sense this is to be expected because whatever is meant here by the respondents, the conditions are severe enough for the individual to be reported in the interview as a residual category. Plausibly, the

lowest disability weights are found for cataracts and high cholesterol levels. Among the measurements, not completing the grip strength test and low walking speed have the highest disability weights. For comparative reasons, I also report disability weights normalised by the difference between the first and the fourth (average) threshold value (CR-disability weights). Overall, the results prove to be similar to those obtained in earlier studies [2,20].

<about here Figure 2>

Let us now turn to the health index values computed using the disability weights described in Table 3. How is it distributed across countries? Figure 2 shows the 25th, 50th and 75th percentile of the age-sex standardised health index distribution by country. The vertical lines at .78, .88, and .95 indicate the (unweighted) average percentiles, so that values to the left of the line indicate that a country's value is below that of the "average" country, values to the right indicate it is above average. The countries are sorted by median health, with the healthiest country (Switzerland) on top and the least healthy country (Spain) at the bottom of the graph. Health inequality (measured by interquartile range) is largest in Spain (0.198) and smallest in Switzerland (0.145).

4. Cross-country differences in response styles

The model used to compute the health index is set up in a way that a value of one is equivalent to the absence of any impairment: no single chronic disease reported, grip strength and walking speed above certain limits, and BMI in the normal range. It is instructive to look first at the proportion of respondents in each country who have a

value of one and the level of health they self-report (Table 4): Switzerland is the country with the largest proportion of respondents in perfect health (12.2%). Sweden (8.9%) and Denmark (8.5%) drop to 4th and 6th rank, respectively. The bottom of the table is filled with the three Mediterranean countries Greece, Italy, and Spain. Now let us consider how these respondents rated their overall health. If everybody really was in perfect health, we would expect 100% of them to report excellent health. This is obviously not the case. Very few respondents said they are in fair or poor health, but the rest of the distribution seems to be centred around "very good" health. The set of variables (diseases and measurements) selected to define perfect health is rather parsimonious. Unobserved health problems might lead respondents to report worse than excellent health, so that some proportion of respondents reporting less than excellent health was to be expected. However, conditional self-rated health distributions vary a lot across countries. Again, the extremes in terms of self-rated health are found in Denmark and Sweden (with more than 40% reporting excellent health) and Italy, France, and Spain (with more than 40% reporting only good health). To account for such large differences, one needs to assume that either unobserved health problems translate differently into self-reported health in different countries or that even perfect health is not reported as the same level in the different countries.

<about here Table 4>

Going back to the full sample, the next step of the analysis is to compute country-specific threshold values for health self-assessments. These thresholds indicate how healthy respondents must be in order to state they are, say, in "good" rather than in "fair" health. To be precise, we are searching for the health index values that need to be exceeded in order to move self-reported health up by one notch. Of course, this value

can differ between individuals as much as it differs between countries, but we are interested in country averages. The approach chosen in the following is thus to compute country-specific threshold values as the exact quantiles of the country-specific health index distribution that correspond to the proportion of respondents that report up to a specific health level. If, for example, 50% of all respondents in a country reported to be in "good" or better health, the health index threshold between "fair" and "good" would be computed as the median of the country-specific health index distribution.

<about here Figure 3>

The results of the calculation are illustrated in Figure 3. Exact numbers are given in the Appendix. Again, the vertical lines indicate the (unweighted) average thresholds, and countries are ranked according to their computed fair-to-good threshold. According to my calculations, Swedish respondents have the lowest threshold between fair and good. In fact, there may be a specific reason for this particular Swedish threshold being so far below that of the other countries (including Denmark). "Fair" and "Poor" are translated as "Ganska dålig" and "Dålig", whereas all other countries have used completely different wordings for each of the two categories (see Table A1 in the Appendix). In any case, conditional on the estimated health index, Swedes are more likely to report good or better health than respondents in all other SHARE countries. The other extreme is represented by German respondents. Conditional on their health index, they are least likely to report fair or better health.

5. Self-reported and adjusted health levels

Given the health index and the reporting thresholds, it is straightforward to compute adjusted distributions of self-reported health. We simply need to use the same (counterfactual) thresholds for each respondent. This could be some specific country's values or the (unweighted) average across all countries. Here, I use the all country average, that is I compute which health level a respondent would report, given his or her health index, if he or she behaved like the average SHARE respondent. Specifically, each respondent whose health index is in the interval between 0 and .62 (the first average threshold) is assigned to poor health, if between .62 and .8 (between the first and second thresholds), the respondent is assigned to fair health, and so on.

<about here Figure 4>

Figure 4 repeats Figure 1, showing the distribution of self-reported health levels across countries. However, this time, it is showing the adjusted distribution, computed as explained above. Overall, cross-country differences have become indeed smaller. Former top countries Denmark and Sweden have moved to the middle ranks, and especially Germany has gained a few positions. The unadjusted percentage of Danes that claim to be in excellent health was 19.8%, compared to only 4.6% in Germany. The adjusted figures are 9.2% versus 9.5%. The proportion of Danes and Germans in "excellent" health is thus quite similar and the differences appear to be more realistic. Still, also after correcting for possible response effects, important health differences remain across countries (as already shown in Figure 3). For instance, the adjusted proportion of Spanish respondents in excellent health is 5.4% and thus only about half as large as in Denmark or Germany. The countries with the largest differences between self-reported and adjusted health level are Germany and Denmark. German respondents

systematically undervalue their health, compared to the SHARE average, whereas Danish respondents systematically overvalue their health. Although there are huge differences in the distributions of self-reported health, adjusted health is much less different between the two countries.

<about here Figure 5>

Figure 5 summarises the message of this paper. It compares the proportion of respondents who actually say they are in very good or excellent health (i.e. country-specific reporting styles) with the proportion of respondents who would say they are in very good or excellent health if they behaved like the average SHARE respondent. In other words, distances in the horizontal direction show cross-national differences in subjective health and distances in the vertical direction show differences in "true" health, as measured by the set of indicators described above. Horizontal and vertical distances to the 45° line indicate the amount to which respondents from different countries over- and underrate their own health (compared to the SHARE average). Respondents in countries located north-west of the 45° line underrate their health and respondents in countries located south-east of the 45° line overrate their health. Again, in the light of the analyses above, it is no surprise to note that Danes and Swedes tend to largely overrate their health, whereas Germans largely underrate their health. Austrians and Greeks show very little bias compared to the average and the rest of countries underrates health somewhat.

6. A simple policy example

Let us look at the implications of the above adjustment for different reporting styles in a simple policy example. One of the major strengths of cross-country data is the possibility to exploit international differences in institutions for policy analyses. Let us assume that we are interested in the relationship between health care expenditures and some simple but comprehensive measure of current health. I start the analysis by looking at the relationship between health care expenditures in 2003 (as percentage of GDP – the data are taken from the OECD) and the proportion of elderly who are in very good or excellent self-assessed health (see left Panel in Figure 6). It appears as if there is no clear relationship between health expenditures and health outcomes. It might be positive, but very weakly. The picture changes if we consider our corrected self-reported health measure (see right Panel in Figure 6). The relationship between expenditures and health becomes positive. Linear regression analysis suggests that a one percentage point increase in health care expenditures is associated with a (statistically significant) 2.6 percentage point increase in the proportion of healthy respondents. This result is robust in the sense that dropping any single (supposedly influential) country from the analysis does not change our finding that health care expenditures are more positively related to a health measure that is adjusted for differences in reporting styles. Of course, this simple example cannot replace a full-blown policy analysis, and it clearly cannot tell us whether 1% of GDP are well spent when it increases the proportion of healthy elderly by 2.6 percentage points. However, it shows that a correction for cross-national differences in reporting styles can affect results substantially.

7. Oaxaca-Blinder Decomposition of Self-Reported Health

The crucial assumption maintained in the earlier sections of this paper is that disability weights are the same for each individual in each country. It is easily possible to test this assumption statistically by running health regressions separately by country and looking whether regression parameters differ significantly between countries. In fact, as it turns out, at least some of the coefficients seem to differ quite substantially. If one allows for differences in disability weights, it is natural to think of decomposing observed differences in self-reported health in three parts: differences due to differences in the prevalence of conditions (a *prevalence* effect), differences due to differences in the effect of conditions on health (a *severity* effect), and a residual difference that contains other unspecified effects including reporting bias [13]. This decomposition follows the Oaxaca-Blinder approach well-known in the labour economics literature (non-parametric decomposition by way of re-weighting [23] yields very similar results).

For simplicity, let us assume that self-reported health is a binary variable *y* that indicates "very good" or better health. The first step of the decomposition is to run country-by-country regressions of self-assessed health on the set of conditions and measurements used in the previous sections to obtain country-specific disability weights. The difference in the proportion of respondents in very good or excellent health in country *k* and the overall proportion can then be decomposed as follows:

$$E(y_k) - E(y_{all}) = \underbrace{(\overline{X}_k - \overline{X}_{all})\hat{\beta}_k}_{prevalence} + \underbrace{\overline{X}_{all}(\hat{\beta}_k - \hat{\beta}_{all})}_{severity} + \underbrace{\hat{\alpha}_k - \hat{\alpha}_{all}}_{other} \qquad (2)$$

where $\overline{X}_k$ is a vector that contains the prevalence of all conditions in country *k*, $\hat{\beta}_k$ is a vector of estimated disability weights, and $\hat{\alpha}_k$ is the proportion of respondents in very good health who report no diagnosed conditions (the regression intercept, also see Table 4 above).

<about here Table 5>

The results of the decomposition into prevalence, severity, and residual effects for the 10 countries are shown in Table 5. By construction, the order of countries is very much the same as in the preceding analyses. In almost all countries, the estimated severity effects have signs opposite to those of the prevalence effect. In other words: in countries with a relatively large proportion of respondents reporting diagnosed conditions and with unfavourable measurements, the estimated disability weights tend to be relatively small. Across countries, more common conditions are associated with lower disability weights. Both effects thus tend to cancel each other out. Therefore, by far the largest part of the difference between the percentage of respondents reporting very good or excellent health in each country and the overall percentage is attributed to "other causes", for instance country-specific reporting styles. Note that the importance of other causes is identified by the cross-country differences in self-reported health among respondents who are factually classified as being in "perfect" health.

In the preceding sections, I have distinguished cross-national differences in self-assessed health due to differences in prevalences and other sources, whereas the Oaxaca-Blinder-style decomposition further isolates from these other sources differences due to differences in disability weights. However, the estimated differences in disability weights or severity effects are hard to interpret. Earlier research has shown

cross-national variation also in expert judgements on disability weights [24]. However, whether and how to compare health cross-nationally if judgements differ is an unsolved ethical question [25]. A related question that should be addressed by future research is why disability weights are different. For instance, are there any systematic relationships with the nature of national health care systems? While good preventive care can decrease the prevalence of at least some of the observed conditions, good health care can attenuate the adverse effects of conditions (not only) on self-reported general health. The results shown in this section suggest that such explanations are currently hard to substantiate. For instance, one referee asked whether the fact that Denmark and Sweden lose ranks after adjusting for response styles could be due to their health systems having a beneficial effect on well-being and thus self-rated overall health. Table 5 actually suggests the opposite. The severity effect assumes relatively large negative effects in those two countries, meaning that given conditions tend to have larger effects on health self-ratings than in the other countries

8. Summary and Conclusion

This paper looks at differences in self-reported health across countries using new data from the Survey of Health, Ageing and Retirement in Europe (SHARE). The specific objective of this paper is to correct observed differences in self-reports for the potential bias caused by cross-cultural differences in reporting styles. In the SHARE data, self-reported general health shows large cross-country variations. According to their self-reports, the healthiest respondents live in the Scandinavian countries and the least healthy in Southern Europe. However, these differences are only partly reflected by differences in true health – as measured by the prevalence of chronic conditions and

objective health measures such as grip strength, walking speed, and BMI. The remaining part of the cross-country variation in self-reported health must probably be attributed to differences in reporting styles. The detailed health data available in SHARE allow to compute health measures that are purged of reporting styles. Such comparable measures are a necessity in cross-national, multidisciplinary analyses. In fact, if differences in reporting styles are taken into account, cross-country variations in general health are greatly reduced but certainly not eliminated, and the order of countries in terms of their population health changes substantially.

References

[1] Idler EL, Benyamini Y. Self-rated health and mortality: A review of twenty-seven community studies. *J Health Soc Behav* 1997; 38: 21-37.

[2] Groot W. Adaption and scale of reference bias in self-assessments of quality of life. *J Health Econ* 2000; 19: 403-420.

[3] Sen A. Health: perception versus observation. *Br Med J* 2002; 324: 860-861.

[4] Lindeboom M, Van Doorslaer E. Cut-point shift and index shift in self-reported health. *J Health Econ* 2004; 23: 1083-1099.

[5] Van Doorslaer E, Gerdtham, U-G. Does inequality in self-assessed health predict inequality in survival by income? Evidence from Swedish Data. *Soc Sci Med* 2003; 57: 1621-1629.

[6] Sadana R, Mathers, CD, Lopez, AD, Murray, CJL, Moesgaard Iburg, K. Comparative analyses of more than 50 household surveys on health status. Ch. 8.1. in Murray, CJL, Solomon, JA, Mathers, CD, Lopez, AD (eds). *Summary measures of population health. Concepts, ethics, measurements and applications.* Geneva, WHO, 2002

[7] National Research Council. *Preparing for an Aging World: the Case for Cross-National Research.* National Academy Press: Washington, DC, 2001.

[8] Crossley T, Kennedy S. The reliability of self-assessed health status. *J Health Econ* 2002; 21: 643-658.

[9] Greene, WH. Econometric Analyis, 5th ed. Prentice Hall: Englewood Cliffs, NJ 2003.

[10] Börsch-Supan A, Brugiavini A, Jürges H, Mackenbach J, Siegrist J, Weber, G, (eds). *Health, Ageing and Retirement in Europe – First Results from the Survey of Health, Ageing and Retirement in Europe.* MEA: Mannheim 2005.

[11] Börsch-Supan A, Jürges H (eds). *The Survey of Health, Ageing and Retirement in Europe – Methodology.* MEA: Mannheim 2005.

[12] Human Mortality Database (at www.mortality.org).

[13] Case A, Paxson C. Sex differences in morbidity and mortality. *Demography* 2005; 42: 189-214.

[14] Mackenbach J, Looman, C, van der Meer, J. Differences in the Misreporting of Chronic Conditions, by Level of Education: The Effect of Inequalities in Prevalence Rates. *Am J Public Health* 1996; 86: 706-711.

[15] Baker M, Stabile M, Deri C. What do self-reported, objective, measures of health measure*? J Hum Resour* 2004; 39: 1067-1093.

[16] Frederiksen H, Gaist D, Petersen HC, Hjelmborg J, McGue M, Vaupel JW, Christensen K. Hand grip strength: a phenotype suitable for identifying genetic variants affecting mid- and late-life physical functioning. *Genet Epidemiol* 2002; 23: 110-22.

[17] Rantanen T, Masaki K, Foley D, Izmirlian G, White L, Guralnik JM. Grip strength changes over 27 years in Japanese-American men. *J Appl Physiol* 1998; 85: 2047-2053.

[18] Nicholas S, Huppert FA, McWilliams B, Melzer D. Physical and cognitive function. In: *Health, wealth and lifestyles of the older population in England: The 2002 English Longitudinal Study of Ageing*, Marmot M, Banks J, Blundell R, Lessof C, Nazroo J (eds). IFS: London 2003.

[19] Torrance GW. Measurement of health state utilities for economic appraisal. A review. *J Health Econ* 1986; 5: 1-30.

[20] Cutler, DM, Richardson, E. Measuring the Health of the U.S. population. *Brookings Pap Econ Act: Microeconomics* 1997: 217-271.

[21] Terza, JV. Ordinal Probit: a generalisation. *Commun Statist Theor Meth* 1985; 14: 1-11.

[22] King G, Murray CJ, Salomon JA, Tandon, A. Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research. *Am Polit Sci Rev* 2004; 98: 191-207.

[23] Barsky R, Bound J, Charles K, Lupton J. Accounting for the black-white wealth gap: a nonparametric approach. *J Am Stat Assoc* 2002; 97: 663-673.

[24] Üstün TB, Rehm J, Chatterji, S et al. Multiple-informant ranking of the disabling effects of different health conditions in 14 countries. *Lancet* 1999; 354: 111-115.

[25] Sommerfeld J, Baltussen R, Metz L, Sanon M, Sauerborn R. Determinants of variance in health state valuations. Ch. 11.1. in Murray, CJL, Solomon, JA, Mathers, CD, and Lopez, AD (eds.*) Summary measures of population health. Concepts, ethics, measurements and applications*. Geneva, WHO, 2002.

Table 1: **Sample size, by country, sex and age group**

Country	Total	Sex		Age Group		
		Men (%)	Women (%)	50-64 (%)	65-79 (%)	80+ (%)
Austria	1,938	42.0	58.0	51.8	39.1	9.1
Germany	2,946	46.6	53.4	53.4	39.6	7.0
Sweden	3,010	47.1	52.9	53.0	36.5	10.5
Netherlands	2,878	47.1	52.9	59.2	32.8	8.0
Spain	2,373	42.1	57.9	46.0	41.1	12.9
Italy	2,506	44.9	55.1	53.4	39.5	7.0
France	1,748	44.9	55.1	53.1	36.4	10.5
Denmark	1,637	47.0	53.0	56.8	32.2	11.1
Greece	1,980	45.4	54.6	52.3	37.0	10.7
Switzerland	956	47.7	52.3	53.1	36.1	10.8
Total	21,972	45.5	54.5	53.3	37.2	9.5

Source: SHARE release 1

Table 2:   **Prevalence of chronic conditions and physical health measures (age-sex standardised; by country)**

	Austria	Germany	Sweden	Nether-lands	Spain	Italy	France	Denmark	Greece	Switzer-land	Total
Heart attack / heart failure	9.4	11.7	15.7	11.5	11.1	10.3	12.8	8.8	11.6	7.2	11.5
High blood pressure	30.0	35.7	28.8	25.7	32.4	35.7	29.1	29.6	36.6	26.3	31.5
High blood cholesterol	15.6	18.0	16.8	14.9	23.5	19.2	23.4	15.3	20.8	13.0	18.4
Stroke or cerebral vascular disease	4.1	4.1	4.4	4.5	2.0	3.2	3.2	5.3	3.7	2.4	3.8
Diabetes or high blood sugar	8.5	11.0	8.4	8.1	14.5	11.1	9.0	7.6	8.5	5.9	9.6
Chronic lung disease	3.1	5.0	2.8	6.6	5.3	7.3	5.5	7.5	3.6	3.2	5.0
Asthma	4.9	3.2	7.4	4.0	3.9	5.2	4.5	7.9	3.4	3.4	4.8
Arthritis	10.6	12.2	9.7	9.7	27.8	31.1	30.2	26.3	17.5	11.6	18.3
Osteoporosis	8.5	7.4	3.1	7.6	8.7	10.6	6.0	3.3	10.8	6.8	7.3
Cancer	3.4	6.2	7.3	6.3	3.4	4.6	5.9	7.9	2.1	5.2	5.3
Stomach or duodenal ulcer	5.7	5.8	4.8	5.1	6.2	6.1	3.6	6.2	7.7	2.3	5.5
Parkinson disease	0.6	0.5	0.5	0.5	0.9	0.6	0.5	0.9	0.9	0.5	0.6
Cataracts	6.3	6.7	10.2	7.7	10.6	6.1	6.7	11.3	7.5	7.3	8.1
Hip fracture or femoral fracture	1.1	1.7	2.8	2.1	2.0	1.6	1.2	2.3	2.3	1.1	1.9
Other	10.3	19.1	28.0	17.3	24.8	14.1	13.0	19.2	8.9	11.8	17.6
Ever treated for depression	7.9	10.7	15.1	16.0	17.7	12.2	21.8	14.5	3.5	13.1	13.4
Low grip strength	20.9	23.8	29.5	28.1	35.9	33.0	26.9	28.9	30.4	26.1	28.4
No grip strength	13.2	8.0	5.2	5.2	5.5	9.7	7.1	2.9	10.5	4.1	7.2
Low walking speed (<0.4m/s)	13.9	14.3	10.0	8.7	13.7	15.3	8.4	8.1	13.4	6.5	11.5
BMI<20[a)] (underweight)	3.1	2.8	4.5	3.3	2.1	3.3	6.0	6.2	2.0	6.7	3.7
BMI 25 to 30 (overweight)	42.7	44.0	40.6	42.4	45.4	43.1	38.1	38.2	47.4	37.1	42.5
BMI 30+ (obese)	18.9	16.9	13.5	14.8	24.0	17.5	15.3	13.4	19.9	13.0	16.9

Source: SHARE release 1; [a)] BMI = Body mass index (weight in kg / squared height in m)

Table 3: **Ordered probit and generalised ordered probit regressions of self-assessed health on health indicators, and implied disability weights**

	Ordered probit	Generalised ordered probit	Implied disability weight[a]	CR-disability weight[b]
Heart attack or other heart problems	0.569**	0.578**	0.098	0.163
	(0.024)	(0.025)		
High blood pressure	0.262**	0.266**	0.045	0.075
	(0.017)	(0.017)		
High blood cholesterol	0.128**	0.130**	0.022	0.037
	(0.020)	(0.020)		
Stroke or cerebral vascular disease	0.707**	0.699**	0.119	0.198
	(0.041)	(0.041)		
Diabetes	0.452**	0.459**	0.078	0.130
	(0.026)	(0.026)		
Chronic lung disease	0.580**	0.571**	0.097	0.161
	(0.036)	(0.036)		
Asthma	0.319**	0.319**	0.054	0.090
	(0.036)	(0.036)		
Arthritis or rheumatism	0.544**	0.545**	0.093	0.154
	(0.021)	(0.021)		
Osteoporosis	0.436**	0.442**	0.075	0.125
	(0.030)	(0.030)		
Cancer or malignant tumour	0.525**	0.527**	0.089	0.149
	(0.033)	(0.033)		
Stomach, duodenal or peptic ulcer	0.317**	0.315**	0.053	0.089
	(0.033)	(0.033)		
Parkinson disease	0.859**	0.857**	0.145	0.242
	(0.102)	(0.101)		
Cataracts	0.077**	0.075**	0.013	0.021
	(0.028)	(0.028)		
Hip or femoral fracture	0.326**	0.326**	0.055	0.092
	(0.056)	(0.056)		
Other condition	0.542**	0.547**	0.093	0.155
	(0.020)	(0.020)		
Ever treated for depression	0.276**	0.278**	0.047	0.079
	(0.022)	(0.022)		
Low grip strength	0.283**	0.285**	0.048	0.080
	(0.017)	(0.018)		
Grip strength test not completed	0.569**	0.578**	0.098	0.163
	(0.032)	(0.032)		
Low walking speed	0.687**	0.692**	0.118	0.196
	(0.027)	(0.027)		
BMI < 20	0.124**	0.116**	0.020	0.033
	(0.041)	(0.041)		
BMI 25 to 30	0.097**	0.099**	0.017	0.028
	(0.017)	(0.017)		
BMI 30+	0.305**	0.309**	0.052	0.087
	(0.023)	(0.023)		
Observations	21,321	21,321		
Ln Likelihood	-25,860.	-25,626.		

Notes – thresholds are listed in Tables A2 and A3 in the Appendix; ** $p<0.01$; Source: SHARE release 1; [a] Disability weights are equal to generalised ordered probit coefficients divided by the difference between the highest and lowest predicted health level; [b]CR = Cutler-Richardson approach (health index standardised by difference between highest and lowest estimated average cutpoint).

Table 4: **Proportion of respondents in "perfect" health (with health index = 1) and their self-reported health levels (age-sex standardised, by country)**

Country	Proportion in "perfect" health...	... self-reporting to be in ... health				
		excellent	very good	good	fair	poor
Switzerland	12.2	27.5	43.3	27.4	1.8	0.0
Austria	9.4	27.2	42.4	26.2	4.2	0.0
Germany	9.2	17.1	44.4	34.1	4.4	0.0
Sweden	8.9	44.7	36.1	19.1	0.0	0.0
Netherlands	8.5	30.2	27.3	39.6	2.6	0.4
Denmark	8.5	46.7	32.2	19.7	1.3	0.0
France	6.7	17.7	34.0	46.5	1.7	0.0
Greece	6.7	25.3	46.8	24.0	3.8	0.0
Italy	6.1	19.9	25.3	48.4	6.4	0.0
Spain	5.0	11.4	41.1	41.3	6.2	0.0

Source: SHARE release 1

Table 5: **Decomposition analysis of the proportion of respondents in excellent and very good self-reported health (age-sex standardised)**

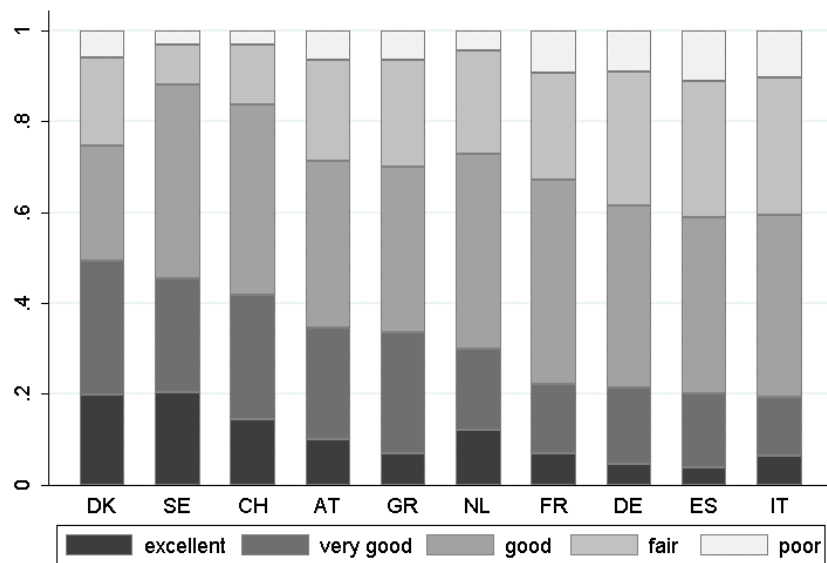| Country	Overall difference to SHARE average	Prevalence effect	Severity effect	Other
Denmark	18.6	0.3	-5.6	23.9
Sweden	13.5	0.7	-6.2	19.0
Switzerland	10.3	5.6	-2.0	6.8
Austria	2.9	2.2	-1.6	2.3
Netherlands	1.8	2.7	-1.6	0.6
Greece	1.7	0.1	-1.1	2.7
France	-8.9	-1.1	6.5	-14.2
Germany	-9.7	0.6	1.6	-11.9
Spain	-10.7	-5.5	8.5	-13.8
Italy	-11.4	-3.2	8.8	-17.1

Source: SHARE release 1

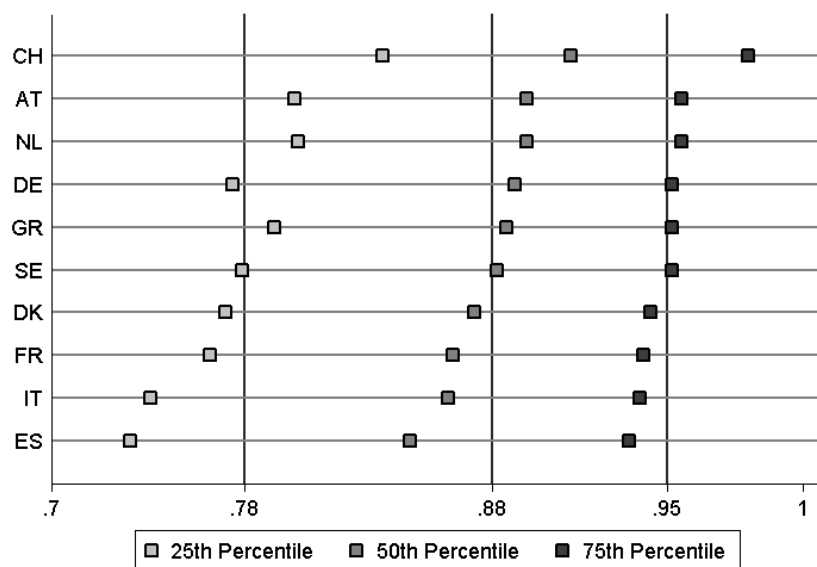**Figure 1:** Self-reported general health, by country

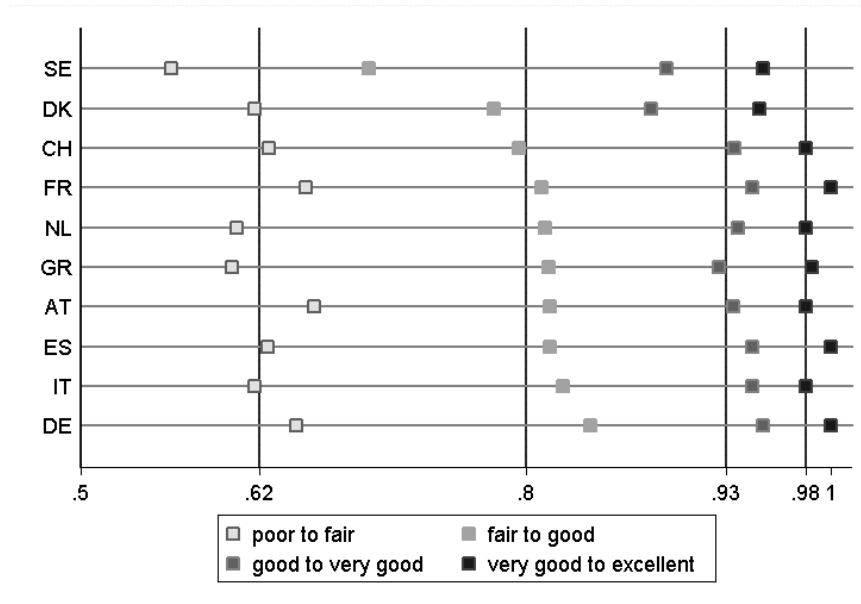**Figure 2:** Distribution of standardised health index, by country
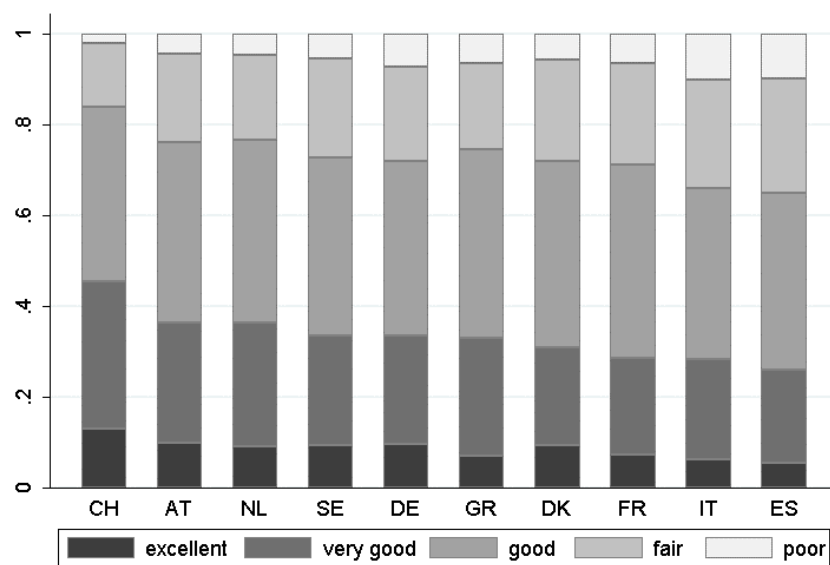
Figure 3: Health index cutpoints, by country

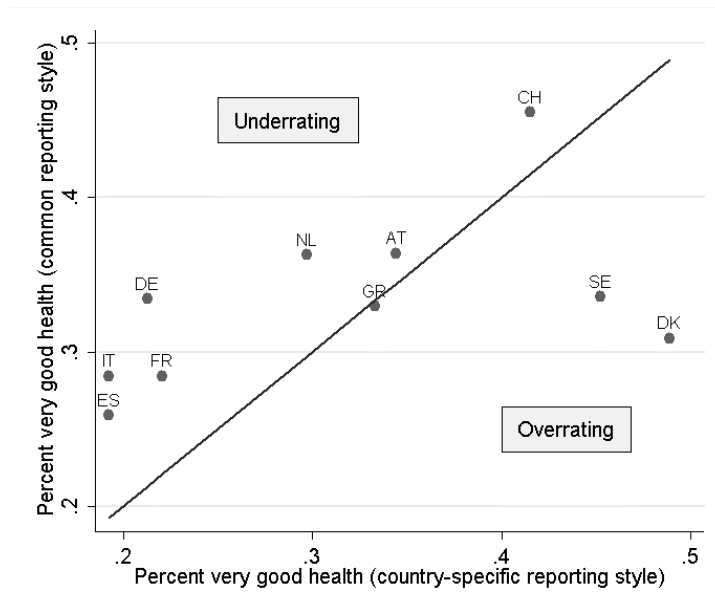**Figure 4:** Adjusted self-reported health levels, by country

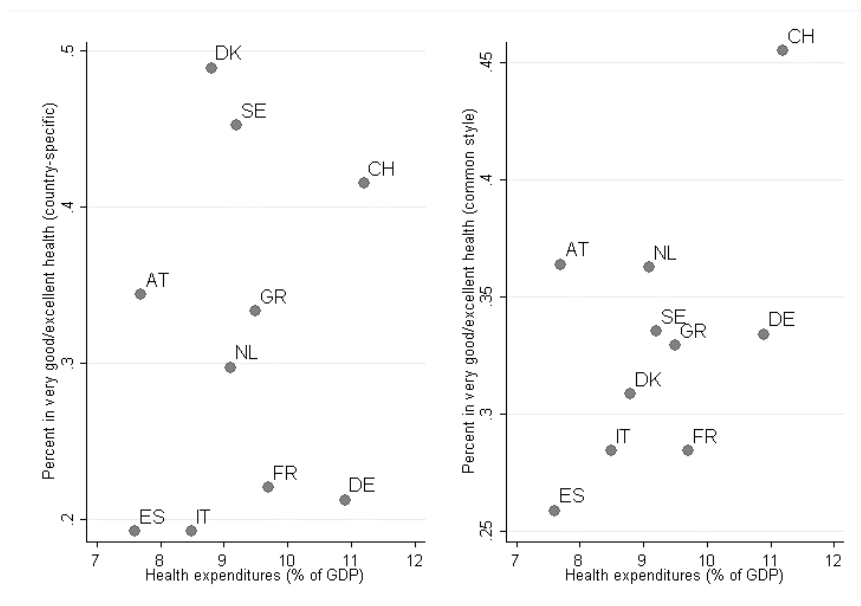**Figure 5:** Self-reported and adjusted health levels

Figure 6: A simple policy example: health expenditures and population health levels

Appendix

Table A1.   **Answer categories for self-assessed health**

Language	Country	Would you say your health is...				
English	Generic	Excellent	Very good	Good	Fair	Poor
German	AT, DE, CH	Ausgezeichnet	Sehr gut	Gut	Mittelmäßig	Schlecht
Spanish	ES	Excelente	Muy buena	Buena	Pasable	Mala
Greek	GR	Αριστη	Πολύ καλή	Καλή	Μέτρια	Κακή
Dutch	NL, BE	Uitstekend	Heel goed	Goed	Redelijk	Slecht
French	FR, BE, CH	Excellente	Très bonne	Bonne	Acceptable	Médiocre
Italian	IT, CH	Ottima	Molto buona	Buona	Discreta	Scadente
Danish	DK	Fremragende	Meget godt	Godt	Nogenlunde	Dårligt
Swedish	SE	Utmärkt	Mycket god	God	Ganska dålig	Dålig

Table A2   **Ordered Probit Threshold Parameters**

Country	Excellent to very good	Very good to good	Good to fair	Fair to poor
Austria	-0.745	0.179	1.532	2.857
Germany	-1.044	-0.120	1.233	2.558
Sweden	-0.127	0.797	2.150	3.475
Netherlands	-0.711	0.213	1.566	2.891
Spain	-0.957	-0.033	1.319	2.645
Italy	-0.969	-0.044	1.308	2.634
France	-0.873	0.051	1.403	2.729
Denmark	-0.290	0.635	1.987	3.312
Greece	-0.732	0.192	1.544	2.870
Switzerland	-0.524	0.401	1.753	3.079

Table A3   **Generalised Ordered Probit Threshold Parameters**

Country	Excellent to very good	Very good to good	Good to fair	Fair to poor
Austria	-0.734	0.292	1.477	2.732
Germany	-1.181	-0.140	1.253	2.676
Sweden	-0.136	0.721	2.373	3.313
Netherlands	-0.587	0.144	1.552	2.999
Spain	-1.128	-0.040	1.329	2.752
Italy	-0.856	-0.101	1.286	2.739
France	-0.840	-0.019	1.495	2.673
Denmark	-0.143	0.848	1.736	2.916
Greece	-0.926	0.322	1.512	2.902
Switzerland	-0.568	0.394	1.849	2.987

Table A4:   **Country-specific cutpoints transforming 0-to-1 health index into self-assessed health categories**

Country	Poor to fair	Fair to good	Good to very good	Very good to excellent
Austria	0.655	0.812	0.935	0.983
Germany	0.643	0.840	0.955	1.000
Sweden	0.560	0.692	0.890	0.955
Netherlands	0.604	0.809	0.938	0.983
Spain	0.625	0.813	0.948	1.000
Italy	0.616	0.821	0.948	0.983
France	0.650	0.807	0.948	1.000
Denmark	0.616	0.776	0.880	0.953
Greece	0.600	0.811	0.925	0.987
Switzerland	0.626	0.792	0.936	0.983